STOCK MARKET PREDICTION USING NATURAL LANGUAGE PROCESSING

Frederick S. M. Herz, etal

ABSTRACT

We present a method of using natural language processing (NLP)
techniques to extract information from online news feeds and then
using the information so extracted to predict changes in stock prices
or volatilities. These predictions can be used to make profitable
trading strategies. More specifically, company names can be recognized
and simple templates describing company actions can be automatically
filled using parsing or pattern matching on words in or near the
sentence containing the company name. These templates can be
clustered into groups which are statistically correlated with changes
in the stock prices.

Our system is composed of two parts:
 a message understanding component
that automatically fills in simple templates

and a statistical correlation component that tests the correlation of
these patterns to increases or decreases in the stock price.

The methods described below can be applied to a broad range of text,
including articles in online newspapers such as the Wall Street
Journal, financial newsletters, radio & TV transcripts and annual

reports. We envision it being used first for newswires such as

Bloomberg, or perhaps the AP Newswire.

In an enhanced embodiment of the system we can further leverage statistical

patterns in Internet usage data and Internet data such as newly released textual

information on Web pages.